Evidence of Clinical Utility: An Unmet Need in Molecular Diagnostics for Patients with Cancer

David R. Parkinson¹, Robert T. McCormack², Susan M. Keating³, Steven I. Gutman⁴, Stanley R. Hamilton⁵, Elizabeth A. Mansfield⁶, Margaret A. Piper⁷, Patricia DeVerka⁸, Felix W. Frueh⁹, J. Milburn Jessup¹⁰, Lisa M. McShane⁶, Sean R. Tunis⁷, Caroline C. Sigman⁷, and Gary J. Kelloff⁹

Abstract
This article defines and describes best practices for the academic and business community to generate evidence of clinical utility for cancer molecular diagnostic assays. Beyond analytical and clinical validation, successful demonstration of clinical utility involves developing sufficient evidence to demonstrate that a diagnostic test results in an improvement in patient outcomes. This discussion is complementary to theoretical frameworks described in previously published guidance and literature reports by the U.S. Food and Drug Administration, Centers for Disease Control and Prevention, Institute of Medicine, and Center for Medical Technology Policy, among others. These reports are comprehensive and specifically clarify appropriate clinical use, adoption, and payer reimbursement for assay manufacturers, as well as Clinical Laboratory Improvement Amendments–certified laboratories, including those that develop assays (laboratory developed tests). Practical criteria and steps for establishing clinical utility are crucial to subsequent decisions for reimbursement without which high-performing molecular diagnostics will have limited availability to patients with cancer and fail to translate scientific advances into high-quality and cost-effective cancer care.

See all articles in this CCR Focus section, "The Precision Medicine Conundrum: Approaches to Companion Diagnostic Co-development."

Clin Cancer Res; 20(6); 1428–44. ©2014 AACR.

Introduction
The widely accepted definition of clinical utility for an assay is that the results of the assay lead to a clinical decision that has been shown with a high level of evidence to improve patient outcomes (1). The demonstration of clinical utility is commonly a two-step process designed to show first that the test provides accurate diagnostic information and second that the diagnostic information, when used in managing patients, given the benefits and harms of the assay, improves health outcomes in a clinically perceivable way compared with alternative management strategies (for at least some number of subjects in the applicable population; refs. 1–6). A test may be considered useful if its results are actionable, driving a treatment decision that leads to a better outcome. These criteria may apply when there is no preexisting test, particularly in comparison to current clinical test results that may not be sufficient to make management decisions. In a broad sense, demonstrating utility may be as simple as showing that a test provides equivalent or increased sensitivity and specificity (leading to equivalent or improved management decisions); is less invasive (incurs less patient harm); is less costly (provides the same benefit with fewer resources); or is more widely or easily available (more likely to be used to make management decisions; refs. 1–8).

Clinical utility has become important as the number of molecular diagnostic tests has grown substantially in number over the last decade. These include tests that specifically measure genetic variability (DNA), gene expression profiles (RNA), or protein expression of biological targets, including assays to identify signaling pathways that contribute to dysregulation of cell proliferation and apoptosis in cancer. For example, the Tufts Evidence-based Practice Center reported that 50 new genetic assays for cancer-related conditions were introduced into clinical use between 2006 and 2011, for breast, colorectal, lung, prostate, and other cancers, resulting in a total of 112 gene-based tests for solid and hematologic tumors (2). The results of molecular assays may be used to assess risk, disease stage, or prognosis; to predict treatment response; and to guide patient therapy (2, 3, 7, 9–15). Examples of some molecular targets and corresponding assays in use today are shown in Fig. 1.

Personalized medicine involving successfully directed targeted therapy depends in large measure on the availability of specific predictive tests to determine which patients express...
Clinical Utility of Molecular Diagnostics for Cancer

Figure 1. Examples of tumor molecular targets for prognostic and predictive biomarker assays commercially available and used in clinical practice. Biomarker assays that are FDA approved, PMA or 510(k), are indicated in bold. Prognostic markers associated with clinical outcome include CEA, EGFR, ER, HER2, PR, and PSA. Predictive biomarkers used to select a targeted therapy include EGFR, EML4-ALK, KRAS, and PIK3CA (7, 11, 13, 28). Abbreviations used: 1p/19q, codeletion of chromosomal arms 1p and 19q; ALL, acute lymphocytic leukemia; AFP, α-fetoprotein (AFP); AML, acute myeloid leukemia; APL, acute promyelocytic leukemia; AR, androgen receptor or nuclear receptor subfamily 3, group C, member 4 (AR); B-CLL; B-cell chronic lymphocytic leukemia; BRAF, v-raf murine sarcoma viral oncogene homolog B1 or proto-oncogene B-Raf (BRAF); BCR-ABL, breakpoint cluster region and Abelson murine leukemia viral oncogene homolog 1 gene fusion; CML, chronic myeloid leukemia; CA125, cancer antigen 125 or mucin 16 (MUC16); CA19-9, cancer antigen 19-9; CEA, carcinoembryonic antigen or glycosyl phosphatidyl inositol (GPI) cell surface anchored glycoprotein; CCRF, continuous human leukemia cell line; CYP2D6, cytochrome P450 2D6; CYP2D6, dihydropyrimidine dehydrogenase (DPYD); EGFR, epidermal growth factor receptor (EGFR); EML4-ALK, echinoderm microtubule-associated protein-like 4 and anaplastic lymphoma kinase fusion (EML4-ALK); ER, estrogen receptor (ER); ESR1, estrogen receptor α (ESR1); FLT3, fms-related tyrosine kinase 3 (FLT3); GIST, gastrointestinal stromal tumor; HE4, human epididymis protein 4 or WAP four-disulfide core domain (HE4); HER2, human epidermal growth factor receptor 2 or receptor tyrosine-protein kinase erbB-2 (HER2); IDH1, isocitrate dehydrogenase 1 (IDH1); IDH2, isocitrate dehydrogenase 2 (IDH2); JAK2, janus kinase 2 (JAK2); KIT, C-CIT, KIT, v-vit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog; KRAS, (KRAS), Kirsten rat sarcoma viral oncogene homolog; LDT, lab-developed Abelson murine leukemia viral oncogene homolog 1 gene fusion; MUC16, membrane glycoprotein; MPD, myelodysplastic disorder; NMP22, nuclear matrix protein 22; NPM1, nucleophosmin 1 (NPM1); NRAS, neuroblastoma RAS viral oncogene homolog (NRAS); NSCLC, non-small cell lung cancer; Oncotype Dx Breast, breast biomarker panel; Oncotype Dx Colon, colon biomarker panel; Oncotype Dx Prostate, prostate biomarker panel; OVA1®, ovariian biomarker panel; PIK3CA, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit α or PI3K p110α (PIK3CA); PML-RARα, promyelocytic leukemia and retinoic acid receptor α fusion (RARα); PR, progesterone receptor (PR); PSA, prostate specific antigen, gamma-semimembranous, kallikrein-3, (KLK3); SCCHN, squamous cell carcinoma of the head and neck; UGT1A1, UDP glucuronosyltransferase 1A1 (UGT1A1).

Two paths for deployment of clinical assays exist in the United States. The first path uses in vitro diagnostic devices (IVD; ref. 18) cleared or approved by the U.S. Food and Drug Administration (FDA; ref. 19). These devices are distributed to Clinical Laboratory Improvement Amendments (CLIA)-certified laboratories, which set up the tests and confirm their ability to perform the test consistent with the standards that have been established by the manufacturer. FDA review generally assures a new assay has established analytical and clinical validity, the latter being defined as the association between the biomarker and the pathophysiologic state or clinical presentation of interest. FDA review occurs before the marketing of a new test and is highly transparent with reviews and review decisions all posted on FDA websites. However, with the sole exception of companion diagnostics (20), the agency does not generally require outcome studies, and so
clinical utility of a new assay is most commonly not established at the time of clearance or approval of a new diagnostic product. The European Union also may be considering a requirement for demonstration of clinical utility for companion diagnostics (21, 22).

The second path uses laboratory-developed tests (LDT) that are not cleared or approved by FDA but are developed and offered by laboratories that are accredited by entities that have deemed status with the Center for Medicare and Medicaid Services (CMS), including the College of American Pathologists (CAP) and the Joint Commission (i.e., laboratories that are CLIA certified; refs. 23, 24). Tests are setup and established at the site of a CLIA-certified laboratory and samples sent to the laboratory for testing. For the second path, there is no premarket evaluation and tests are regulated through spot checks of documentation during the course of a routine laboratory inspection. This documentation is rarely available to the public. Of note, CMS (CLIA) has no authority to mandate that either clinical validity or clinical utility be established for LDTs.

The availability of two paths to market, one without a requirement for establishment of either clinical validity or utility, has remarkably expedited entry of new and cutting edge diagnostics into medical use. For example, the Tufts Evidence-based Practice Center counted molecular tests being performed in laboratories for clinical use in 2010, including non–FDA-approved LDTs, as well as tests using commercial FDA-approved kits, or analyte-specific reagents (ASR; ref. 25). This group found 145 non–FDA-approved LDTs out of 212 tests for solid cancers, and 221 LDTs out of 388 tests for hematopathology. However, whether or not patient interests have been well served remains a question. It is clear that the performance of many new tests now being marketed has not been well served and/or well documented. In addition, there is no assurance that a LDT offered by a given CLIA laboratory is strictly comparable to that offered by a separate CLIA laboratory; reagents, methodologies, and result interpretations may be laboratory specific and in the absence of external review and comparative studies (except where proficiency testing programs are available) potential discrepancies remain unknown (see also ref. 26). Hence, it is not surprising that a number of publications and reviews by groups such as the IOM, CMTP, Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group (EWG), and the ECRI Institute Evidence-based Practice Center under contract to the Agency for Healthcare Research and Quality, have pointed out uneven success in clinical integration of many of the molecular diagnostic assays because of a variety of factors related to both assay performance and clinical evaluation (Table 1; refs. 1, 4, 5, 8, 10, 14, 25, 27–31). Furthermore, independent review has found that many of these assays have not shown sufficient proof of clinical utility (5, 25).

Methodologic and clinical challenges in evaluating clinical utility

A major challenge in establishing performance for new tests is the fact that frequently different tests for the same analyte are poorly standardized. It is not uncommon for an assay of a particular type to be offered using different targets, originating from different sources, perhaps using different methodologies, reagents, and/or computational analysis, all of which can contribute to dissimilar results and hinder successful clinical adoption. Measurement of DNA, RNA, or protein may be directed at the same functional molecular target, but when used to test the clinical outcome may exhibit markedly different analytical and/or clinical sensitivities and specificities. For example, as has been well documented, levels of HER2 (ERBB2) DNA, RNA, and protein are routinely measured in patients with invasive breast cancer to determine prognosis and patients’ candidacy for trastuzumab or other related inhibitor therapy (13, 32–34). A wide variety of assay technologies have been applied to this measurement, including gene amplification using FISH or chromogenic in situ hybridization (CISH), RNA overexpression using quantitative reverse transcription PCR (qRT-PCR), protein product overexpression with immunohistochemistry (IHC) assays, and protein-based assays to measure circulating levels of the external receptor domain. Currently, at least 10 HER2 tests encompassing these various methodologies have been approved by FDA as companion diagnostics (ref. 35; Fig. 1 and Table 2). Results from these methods are reported using different scoring systems (some quantitative, others semiquantitative) often based on subjective, user dependent determinations. Each method carries assay-specific issues (specimen requirements, false negative/positive rates, standardization, and cost) that complicate the decision about which assay to use. More importantly, interlaboratory result comparison/concordance for the same method and between different methods has not been ideal. These issues have led to standardization of these methods over time, due in part to guidelines generated by the American Society of Clinical Oncology (ASCO) and CAP (34, 36–38).

Table 1. Factors contributing to lack of success of molecular diagnostic tests in the clinic

- Method-specific issues
- Availability of multiple assays and methodologies for one target with a lack of material and method standards
- Lack of complete understanding of the clinical setting or scientific knowledge of the disease parameters by the developers
- Insufficient demonstration of analytical and clinical validation before commercialization
- Inadequate clinical trial designs or infrastructure needed to support assay implementation
- Lack of direct evidence that use of the assay leads to improved patient outcomes over other available solutions
- Lack of agreement among stakeholders on definitions and evaluation of evidence of utility

See refs. 1, 4, 5, 8, 10, 14, 25, 27, and 28.
Lack of guidelines for evaluating clinical utility

In addition to method-specific issues, an important obstacle to the clinical success of molecular diagnostic assays, with few exceptions, has been lack of evidence that use of a particular assay will result in a net improvement in the patient's condition. This situation is partially because of lack of clear uniform guidelines for assay developers on generating evidence of associations that are clinically useful in a defined patient population, and partially to the implementation barriers for demonstrating clinical test performance cited above (1, 10, 39–41).

An assay that demonstrates clinical utility by definition improves patient outcomes when used appropriately. An assay with proven clinical utility provides accurate diagnostic information for the patient population, improves patient management, and/or reduces costs. During development of diagnostic assays, data are typically generated on the assay performance for analytical validation and for clinical validation (the association between the marker and the pathophysiological state of interest; Fig. 2). Data are less likely to be generated to demonstrate the association of the test result to clinical decision making affecting patient outcomes and even less likely to demonstrate an impact on costs. This evidence is often collected only through longer term clinical use and post-market commercial studies, which can be expensive, time consuming, and require defined patient populations. It is not surprising that those weighing resources may find these studies unattractive and view them as less critical to commercial success. However, earlier demonstration of utility as part of the assay development process would provide a greater chance of assay adoption. The failure to demonstrate clinical utility is documented in a 2013 report on three commercially available genetic or molecular pathology tests used to identify tissue of origin of cancers in patients with unknown primary sites. These tests were reviewed for analytical and clinical validity, and clinical utility. Although the tests (CancerTypeID, miRview, and PathworkDx) seemed to offer similar clinical accuracy, the authors found insufficient evidence to assess the effects of the tests on treatment decisions and outcomes, i.e., clinical utility (42). Another report published in 2012 evaluated literature for commercially available single nucleotide polymorphism panels to determine risk of prostate cancer (43). The authors concluded that not only was there no evidence available about the clinical utility of the panels, but there was also insufficient evidence to assess analytic and clinical validity. None of these commercially available panels had been evaluated for use in routine clinical settings; and there was no indication that FDA approved the tests.

### Table 2. FDA-approved tests (PMAs) for overexpression and amplification of HER2

<table>
<thead>
<tr>
<th>Device name</th>
<th>Year/manufacturer</th>
<th>Target organ(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONCOR(R) AMPLITECT</td>
<td>1997 (withdrawn 2007)</td>
<td>Breast</td>
</tr>
<tr>
<td>HER/NEU/ERBB2/GENE AMPLIFI</td>
<td>2000</td>
<td>Ventana Medical Systems, Inc.</td>
</tr>
<tr>
<td>Inform Her2/NU (FISH)</td>
<td>2000 (2013)</td>
<td>Breast</td>
</tr>
<tr>
<td>Inform HER2 Dual ISH DNA Probe Cocktail</td>
<td>2011 (2013)</td>
<td>Breast</td>
</tr>
<tr>
<td>InSite HER2/NEU Kit (IHC)</td>
<td>2004 (withdrawn 2006)</td>
<td>Biogenex Laboratories, Inc.</td>
</tr>
<tr>
<td>Bond Oracle Her2 IHC SYSTEM</td>
<td>2012</td>
<td>Leica Biosystems</td>
</tr>
<tr>
<td>HER2 CISH PharmDx Kit</td>
<td>2011 (software 2013)</td>
<td>Dako Denmark A/S</td>
</tr>
<tr>
<td>HER2 FISH PharmDx Kit</td>
<td>2005 (2013)</td>
<td>Dako Denmark A/S</td>
</tr>
</tbody>
</table>

*Information as given on the FDA website (35). Note also, 510(k) approvals for HER2-related diagnostics, including MammaPrint and TargetPrint (Agendia BV) and imaging and software analysis systems can be found in the same place (90).

**Two or more year entries indicate original and latest PMA submissions for the product.
Implementation barriers for adequate assessment of clinical utility also arise from other issues. These can include lack of sufficient standards and reference materials as well as reference methods that make it impossible to address the issue of cross-assay validation (5), lack of appropriate infrastructure and training to support correct implementation and interpretation of tests, and lack of clarity in the clinical application of results. A high-risk commercialization pathway for developers of diagnostics remains one of the major barriers to assessment of clinical utility. This is because of many factors, including risk of obsolescence from rapidly developing technologies, along with ease of access to the market for these new complex technologies. These also include unclear reimbursement criteria from payers involving nonpayment or reimbursement amounts inadequate to recover test development and/or ongoing costs of performance. In aggregate, these commercial barriers have often led to inadequate assessment of the clinical utility of diagnostic tests. However, this landscape might markedly change with new reimbursement requirements by third-party payers calling for a demonstration of clinical utility before payment. If this concept of pay for performance is sustained and maintained, there will be significant incentives to develop new assays with patient outcomes in mind. A likely consequence is the introduction of fewer but better credentialed tests in the future (31).

Recent reports and next steps

The Foundation for the National Institutes of Health (FNIH) Biomarkers Consortium is a public–private biomedical research partnership that promotes development and qualification of promising biomarkers for prevention, early detection, diagnosis, and treatment of disease. Consortium partners include patient advocates, clinical researchers, and representatives from the therapeutic and diagnostic industries, NIH/NCI, and FDA. The Cancer Steering Committee (CSC) Clinical Utility Working Group of the Biomarkers Consortium determined that it would be beneficial to provide practical definitions of clinical utility and guidelines for developers and to outline the information needed by payers to support reimbursement for molecular diagnostic assays used in cancer therapy.

Recent work on policy and frameworks to gather and report evidence coming from IOM (1) and CMTP (5) as well as other groups, including the National Cancer Institute (NCI), Division of Cancer Treatment and Diagnosis (DCTD; ref. 4), CDC, the Genomics Resources Technology and Evaluation Center (TEC-BCBSA), the Center for Health Research (Kaiser Permanente Northwest), and the EWG, have all generated suggestions for implementation (Table 3). There is consensus that if biomarkers are to be translated into clinical practice more effectively, planning for demonstration of clinical utility and addressing the evidence gap should begin earlier in assay development—at least at the clinical study planning stage. This approach would mean incorporating elements in studies to build evidence focused on outcomes when constructing premarket product or clinical development protocols and planning clinical validation trials instead of waiting until postcommercialization.

In alignment with this mission, the CMTP has recently developed an Effectiveness Guidance Document containing recommendations for improving the evidence base for molecular diagnostics from the perspective of end-user and policy decision makers such as clinicians, patients, and payers. This project identified gaps and recommended specific steps that can be taken to improve the evidence base (5). Among the nonmethodological recommendations, the report notes that public–private partnerships such as the Biomarkers Consortium can help advance evidence generation of clinical utility for molecular diagnostic tests with expanded access to sources of funding and data that are needed to support well-designed studies that produce evidence of clinical utility. These studies would be able to leverage the infrastructure and tools that accompany the cancer clinical trials enterprise.
Given the recommendations generated to date, it is appropriate to move toward stakeholder collaboration to translate these guidelines into practical realities that actually implement and fund studies to demonstrate clinical utility for diagnostic assays. This document contains a rational guide and evidentiary standards for demonstrating clinical utility for predictive/prognostic diagnostics for patients with cancer. The article is not a review of previous studies, discussion of possible regulatory review, or description of coverage policy issues, but is directed at how to practically demonstrate clinical utility. It promotes practices consistent with FDA and CMS assessments of the level of evidence required, and the minimum level of clinical validation needed for demonstration of clinical utility.

Current Perspectives on Components of Clinical Utility

As illustrated in Fig. 2, analytical and clinical validation must precede demonstration of an assay’s clinical utility. Analytical validation demonstrates assay performance in the laboratory setting. As defined above, clinical validation is the association between the test result and a pathophysiological state or clinical presentation of interest. This is commonly described in terms of diagnostic accuracy, a measure of how well a test detects or predicts the clinical presentation. Test performance is commonly reported as clinical sensitivity and specificity and/or positive or negative predictive value (3, 4, 14, 27, 29, 44–47). Evaluation of clinical validity by test performance in relevant clinical samples will uncover the variability and significance of a biomarker in subjects with and without the specific disease associated with the molecular target. Practical guidances, discussions, and checklists for evaluating analytical and clinical validity of clinical molecular can be found in the references cited above, see for example refs. 3, 4, 25, 44, and in the ACCE Model list of 44 Targeted Questions aimed at a Comprehensive Review of Genetic Tests (48).

A number of publications discussing clinical validity and utility for cancer molecular diagnostic assays provide detailed tools for evaluation of assay utility, such as those described in Table 4 (see, e.g., ref. 47). A number of these reviews have been sponsored by organizations that have also developed databases of genomic tests and relevant information. Other groups have discussed the advantages and disadvantages of various clinical trial designs for biomarker clinical utility as described in Table 5 (11, 40, 49). All these authors note that the choice of design should be based on the level of confidence that the marker will predict outcome.

Table 3. Previous comprehensive reviews on clinical utility

<table>
<thead>
<tr>
<th>Authors/organization</th>
<th>Title</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Olson S, Berger AC: Institute of Medicine of the National Academies</td>
<td>Genome-based Diagnostics: Clarifying Pathways to Clinical Use: Workshop Summary</td>
<td>(1)</td>
</tr>
<tr>
<td>Woodcock, J. Center for Drug Evaluation and Research, US Food and Drug Administration</td>
<td>Assessing the Clinical Utility of Diagnostics Used in Drug Therapy</td>
<td>(29)</td>
</tr>
<tr>
<td>Schilsky RL, Doroshow JH, Leblanc M, et al.: National Cancer Institute (NCI), Division of Cancer Treatment and Diagnosis (DCTD)</td>
<td>Development and Use of Integral Assays in Clinical Trials</td>
<td>(4)</td>
</tr>
<tr>
<td>The Genomics Resources Technology Resource Center, Centers for Disease Control P; Public Health Genomics: Genetic Testing: Lin, JS, Thompson M, Goddard KAB, et al.: Center for Health Research Kaiser Permanente Northwest, Blue Cross Blue Shield Association (TEC-BCBSA), Department of Primary Care Health Sciences, University of Oxford</td>
<td>ACCE Model System for Collecting, Analyzing and Disseminating Information on Genetic Tests</td>
<td>(91)</td>
</tr>
<tr>
<td>• The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: Methods of the EGAPP Working Group • Outcomes of Interest in Evidence-based Evaluations of Genetic Tests • Improving the Efficiency and Relevance of Evidence-based Recommendations in the Era of Whole-Genome Sequencing: An EGAPP Methods Update</td>
<td></td>
<td>(6, 92, 93)</td>
</tr>
</tbody>
</table>
Table 4. Resources for evaluating clinical utility

<table>
<thead>
<tr>
<th>Resource</th>
<th>Tools</th>
<th>References</th>
</tr>
</thead>
</table>
| Tumor Marker Utility Grading System (TMUGS): a Framework to Evaluate Clinical Utility of Tumor Markers | • Guidelines, TMUGS, to consider when evaluating clinical utility of tumor markers  
• Factors that can impact the associated assay method  
• Definition of levels of evidence for grading utility (level 1 (most definitive)–V) and update | (8, 56) |
| Task Force Report: Evaluating the Clinical Utility of Tumor Markers in Oncology  
National Comprehensive Center (NCCN) | • Discussion of TMUGS levels of evidence  
• Level of evidence ratings for selected glioma, breast, and colon cancer tumor markers | (7) |
| Quality, Regulation and Clinical Utility of Laboratory-Developed Molecular Tests | • Report and database on website of 630 laboratory performed molecular FDA-approved and non-FDA-approved tests, kits or ASRs performed in laboratories for clinical in 2010 use in oncology for solid and hematological metastasis in the Medicare applicable age group | (25) |
| Technology Assessment Report | • Analysis of processes developed for examining clinical validity and utility of molecular tests | |
| Tufts Evidence-based Practice Center | • Points to 7 reviews of clinical utility of molecular tests for cancer diagnosis or treatment monitoring reporting patient-centered health outcomes, including survival, disease recurrence, and treatment | |
| Update on Horizon Scans of Genetic Tests Currently Available for Clinical Use in Cancers | • Report on genetic and genomic (DNA, RNA, and protein) tests available for clinical practice in 2011 for assessing solid and hematological cancer in the Medicare applicable age group, for diagnostic, prognostic, predictive, monitoring, and recurrence detection uses; database, GeneTestTracker, of 121 tests on website | (2) |
| Technology Assessment Report | • Based on published, gray literature (manufacturers and other commercial and non-commercial, organizational websites (e.g., genetests.org, the Association for Molecular Pathology, Pharm GKB)  
• Contains test descriptions, purpose, clinical use, availability, and FDA approval status | |
| Tufts Evidence-based Practice Center | • Model system for analyzing and updating existing data on the safety and effectiveness of DNA-based genetic tests and testing algorithms (diagnostic assays not covered) | (91) |
| ACCE Model System for Collecting, Analyzing and Disseminating Information on Genetic Tests | • Evaluations of genomic assays posted on the website  
• Evaluations based on analytical validity, clinical validity, clinical utility, and associated ethical, legal and social implications  
• Evaluation criteria posted on website; address the natural history of the disorder, impact of the test on patient care, quality assurance, financial costs and benefits, facilities, personnel, and educational materials, and long-term performance evaluation | |
| CDC Office of Public Health Genomics and the Foundation for Blood Research (FBR) | • Methods for evaluation of evidence of clinical validity and utility of genetic tests; expanded to encompass outcomes; added whole-genome sequencing | (6, 92–96) |
| EGAPP Initiative and Working Group (EWG) | • Evaluation of clinical utility based on intended use and the chain of evidence demonstrating clinical utility; importance of the quality of components, consistency and generalizability of the findings  
• Description of acceptable evidence  
• Grading system for the quality of evidence of components of clinical utility | |
| Genome-Based Diagnostics: Clarifying Pathways to Clinical Use: Workshop Summary | • Assay development roadmap to establish clinical utility from definition of purpose and technical feasibility, development studies, analytical methods to cost-effectiveness studies, and acceptable clinical trial designs | (1, 97) |

(Continued on the following page)
Table 4. Resources for evaluating clinical utility (Cont’d)

<table>
<thead>
<tr>
<th>Resource</th>
<th>Tools</th>
<th>References</th>
</tr>
</thead>
</table>
| Evolution of Translational Omics: Lessons Learned and the Path Forward (Chapter 4) | • Recommendations and processes for assessing clinical utility of "omics"-based tests  
• Defines a “bright line” between definition, validation and lock-down of a test and evaluation for clinical utility |  |
| Institute of Medicine | • Address communications with the FDA about the IDE process and validation requirements |  |
| Evaluation of Clinical Validity and Diagnostic Tests in Adult Oncology | • Recommendations on mechanisms that overcome existing barriers to generate evidence to attain clinical utility, including clinical study designs  
• Importance of the context of the test in therapeutic decision making (patient population identification, biomarker prevalence, and assay validation) | (5) |
| Effectiveness Guidance Document | • Assessment of the impact of the test on health outcomes (measures appropriate and validated for the clinical context, patient’s perspective) |  |
| Center for Medical Technology | • Clinical benefits measures  
• Addresses underlying causes for delay in translation of novel biomarkers into the clinic |  |
| Development and Use of Integral Assays in Clinical Trials | • Analytical, clinical and regulatory issues encountered in developing assays to select and stratify patients, assign treatment, and assess outcomes in clinical trials | (4) |
| Criteria for the use of omics-based predictors in clinical trials | • 30 criteria for the use of "omics"-based predictors in NCI-supported clinical trials  
• Clinical trial design issues | (39, 40) |

Abbreviation: ASRs, analyte specific reagents.

*See Appendix A for the ACCE model list of questions for review of clinical utility for genetic tests.

*See Appendix B for examples of the EGAPP methods for evaluation of clinical utility.

Best Practices for Demonstration of Clinical Utility

Historical perspectives on elements of clinical utility, as described above (1, 4–6, 11), lead to the development of recommendations for best practices to establish clinical utility of diagnostics. Given this, clarity about the intended use and context of use in clinical decision making is critical from the beginning of test development, because the intended use and context should inform the design of appropriate test validation and selection of appropriate trial designs and patient populations. Both intended use and context of use are critical in determining the best practices to use across assay development, project, and commercialization planning to lay the foundation for attainment of clinical utility as outlined in Table 6 and stated below.

Define the intended use of the assay along with the clinical context of the test early in assay development and project planning

Intended use refers to the specific purpose of the test and should include what is being measured and why. For example, the intended use may be for screening, diagnosis, or prognosis (response, progression, or disease monitoring) for a specific disease condition associated with a specific target(s). An important component of intended use is identification of the populations to be tested. If possible, studies should ultimately be conducted in these selected populations to assure that predictions of performance in real world use will be accurate.

Definition of intended use requires understanding the clinical context (clinical need and any preexisting solutions) to develop an assay that is fit for purpose. Considerations include whether or not there is a preexisting clinical solution/assay in place for the condition, that is, whether or not the test fits an unmet need. If the proposed assay does not address an unmet need, a consideration should be whether or not it offers other advantages over preexisting tests, that is, is the new test more accessible, less expensive, or more convenient. Potential performance characteristics of the new test and those of preexisting tests should be compared, and analytical and clinical validation studies that have been done should be described. Given the potential benefits and harms of each assay, consideration should be given to how the new diagnostic approach will improve patient management strategies in a clinically perceivable and meaningful way and whether or not the new or preexisting assay would be most likely to be effective in real-world clinical and laboratory situations. The fit for purpose concept provides justification for development of a test that may not have adequate performance characteristics (e.g., sensitivity or specificity) in certain applications but fills an unmet medical need, for example, by identifying a subpopulation for which a new therapeutic demonstrates markedly improved outcomes or by identifying a low prevalence population for which a new
therapeutic shows high efficacy. It will also be important to consider any incentives that will affect the adoption of the new test, from ease of test performance outside of the developer’s laboratory, to assay costs and market size. Results of the above analyses will help in developing the correct hypotheses for testing to establish clinical utility and in determining what types of evidence must be generated to make a case for utility. Having this information early in development of a new assay will help sponsors and other stakeholders make informed decisions about how to prioritize use of resources.

Consider regulatory and commercial implications of implementing an LDT versus an FDA-approved IVD

As described above, there are currently two regulatory routes for a new diagnostic test to enter the market:

| Table 5. Clinical trial designs in relation to demonstration of clinical utility a |
|----------------------------------------|--------------------------------------------------|-------------------|
| Trial design                           | Description                                                                                      | Recommendation |
| Prospective-randomized clinical       | All comers marker-stratified design; patients separated into marker positive and negative arms  | Preferred       |
| trial design without enrichment for    | Patients randomized to treatment with drug A or B in each arm                                      |                  |
| marker status.                         | Biomarker evaluated in all patients, but randomization to treatment restricted to marker positive patients | Not preferred |
| Marker enrichment b                    |Marker evaluated in only one arm                                                                   | Not preferred |
| Marker-based strategy (enrichment) b  | Requires:                                                                                       | Acceptable      |
|                                      | • Enough appropriate biopsies (archived tissue from at least two thirds of patients) must be available from the original appropriately powered drug trial to assure that patients included in the biomarker evaluation are representative of patients in the source trial |                  |
|                                      | • Analytical validation of the assay to ensure that results from archived tissues resemble results obtained from real-time tissue collection |                  |
|                                      | • Prespecified analysis plan used in marker study                                                |                  |
|                                      | • Results validated in one or more similarly designed studies using the same assay techniques    |                  |
| Prospective–retrospective analysis of | Patients randomized to treatment with drug A or B in each arm                                      |                  |
| new markers in all patients from a    | Requires:                                                                                       | Acceptable      |
| previous trial of a previously        | • Enough appropriate biopsies (archived tissue from at least two thirds of patients) must be available from the original appropriately powered drug trial to assure that patients included in the biomarker evaluation are representative of patients in the source trial |                  |
| developed therapeutic                 | • Analytical validation of the assay to ensure that results from archived tissues resemble results obtained from real-time tissue collection |                  |
|                                      | • Prespecified analysis plan used in marker study                                                |                  |
|                                      | • Results validated in one or more similarly designed studies using the same assay techniques    |                  |
| A single-arm prospective study        | If sufficient archived tissues are lacking to conduct a prospective–retrospective study, a well-conducted study demonstrating that the companion diagnostic test can distinguish responders from nonresponders would be considered adequate evidence of the clinical utility of the test | Acceptable      |
| (companion diagnostic to an approved  |伴侶診斷沧州to an approved drug with previously established efficacy)                              | Acceptable      |
| drug with previously established      |伴侶診斷沧州Development process should produced sufficient evidence of clinical utility       | Acceptable      |
| efficacy)                             | Modelin伴侶診斷沧州Modeling the clinical utility of a test may take advantage of existing evidence (i.e., from preexisting studies) to form an indirect evidence chain, or a clear surrogate marker linked to a health outcome | Acceptable      |

aDesigns are as outlined in refs. 1, 4, 11, 14, 17, 25, 39, 40, 49, 55, 56, 98, and references therein.

bEnrichment designs are designs that exclude patients with a specific marker status and designs that use standard of care unevenly between study arms. They are not always preferred; these authors note that from a statistical point of view, study designs should be based on the level of confidence that the marker will predict outcome. Given this, enrichment designs are appropriate when preliminary evidence suggests patients with or without the marker profile will not benefit from the treatment. The unselected design is optimal when preliminary evidence for treatment benefit or assay reproducibility is uncertain. A hybrid of the enrichment and unselected designs may be appropriate if ethical issues are involved.
performance in terms of patient outcomes, specifically.

agnostic development, the FDA does assess diagnostic per-
drug to be approved. In this unique case of drug-diag-
approved if it is essential to the safe and effective use of a
diagnostics, FDA requires that a companion diagnostic test be
ent part of clinical validity (e.g., FDA companion
decision making unless those outcomes are an inher-
either routinely looks at clinical outcomes as part of
FDA nor CMS (CLIA) specifically requires this and
in which results are masked and not used in clinical
decision making, clinical users, and third-party payers frequently
have difficulty interpreting and implementing the
recommendations.

In spite of uneven or nonexistent requirements for clinical
outcome studies by FDA and CMS (CLIA), units at CMS
making coverage or reimbursement decisions, regional
Medicare carriers, other third-party payers, and other
stakeholders invested in evidence-based medicine do have
considerable interest in clinical utility. Recent changes in
reimbursement for molecular diagnostics (replacement of
stacking codes with test-specific codes) have reinforced use
of clinical utility as an important ingredient in decision
making about reimbursement of tests. Whether or not a new
test is developed and undergoes FDA review or is offered
without FDA review as an LDT, an important step in
obtaining payment for the test and ensuring market uptake
is to develop evidence that the test does change patient
outcomes.

Although FDA and CMS have initiated a pilot to allow for
joint decision making about regulatory approval/clearance
and payment, in general these two processes are not con-
ected. Sponsors and other champions of new diagnostic
tests should understand this sometimes unappreciated bar-
tier to successful introduction of a new test and establish a
plan for addressing this.

Both IVDs reviewed by the FDA or offered as LDTs
without FDA review are intended for clinical use; hence,
can only be performed in laboratories with CLIA certifica-
tion. As stated above, this certification involves inspections
and proficiency testing through CMS, CAP (50), or other
designated third parties. Thus, early planning to meet
inspection guidelines, which are standard for biomarker
laboratory practice, will benefit efficient development of
any test, particularly an LDT.

In planning for clinical development of a new assay, it
is also important to determine if an Investigational Device
Exemption (IDE) is required while the assay is under
investigation. An IDE is generally not required for studies
in which results are masked and not used in clinical
decision making, but is required when important patient
care decisions are being made based on an investigational
test. An exception is phase I studies where the efficacy and
safety of a new therapeutic is unknown and the FDA has
allowed the use of an assay described in the IND for
patient selection.

The FDA might require an IDE filing to collect trial
safety and effectiveness data depending on the risk of use
of the test, often to support premarketing approval (PMA)
submissions. Use of an investigational device for clinical
management often requires an approved IDE before the
study is initiated, again depending on the risk of the

Table 6. Key steps toward demonstration of clinical utility

- Define the intended use of the new test and associate the
  assay result with a measurable clinical outcome.
- Understand the clinical context of the test in relation to the
  clinical need and disease. Determine what increased benefit
  (e.g., effectiveness vs. efficacy, access, cost, and/or
  performance) a new test would bring to the clinical setting
  over established procedures.
- When evaluating regulatory and business strategies,
  consider the differences and similarities in implementing an
  FDA-approved test (IVD) vs. an LDT test without FDA
  clearance, with CLIA certification.
- Outline what will be needed to demonstrate clinical utility as
  part of assay implementation (e.g., infrastructure, sample
  logistics, user training materials, effectiveness/quality
  measures, and components for long-term performance
  tracking).
- Anticipate the demonstration of clinical utility to support
  regulatory approval and payer clearance, including cost–
  benefit analysis.
- Integrate all the steps needed to demonstrate clinical utility
  into the product development plan, anticipating what
  evidence will be collected pre- and postapproval and
  commercialization.
- Whether developing an LDT to be offered without FDA
  review or an IVD for FDA approval, carefully design clinical
  trials and specify an analysis plan to start generating
  evidence of clinical utility before regulatory approval and
  commercialization; work with partners to obtain access to
  appropriate sample and patient populations.
- Assemble the chain of evidence; study data in context of
  use/patient population of interest, measured outcomes,
  published results, economic impacts, and compare to
  current practices and impacts on physician behavior or
  changes in clinical practice.
device used to direct management, that is, whether or not an incorrect test result or a ‘bad test’ could expose patients to serious harm or death (51). It is helpful to consult with the FDA and consider a pre-IDE meeting if and when appropriate. It may be worthwhile to consider a managed entry scheme or a staged approach to market entry and to work with business partners and FDA to explore nontraditional approaches to test implementation, for example, FDA-CMS parallel review described above, adaptive licensing, and/or performance-based risk-sharing arrangements with a provision of coverage for patients in well-designed clinical trials to gather evidence of clinical utility (5, 52).

Outline what components of assay implementation will be needed to demonstrate clinical utility, including logistical infrastructure

Test developers should consider how effectiveness (uptake; ability to deliver), as opposed to efficacy, will be demonstrated. This will include determining how test performance will be monitored, how safety signals from use of a new diagnostic will be analyzed, and whether or not formal or informal assessments of test use over time will be performed. Assay performance may be studied using carefully constructed clinical trials, using existing or new registries, and/or monitoring carefully selected safety signals in an ordered descriptive manner. An important issue in addressing new test use is ensuring that interested stakeholders (clinical users, patients, others) have a clear understanding of how the test should be used and order or request it appropriately. In some cases, this may be obvious but in others it may be helpful and important to provide educational (training materials, guidelines for use) or regulatory materials. A practical example of a successful train of documentation to establish a new test is illustrated on the Genomic Health website, documenting clinical studies and literature around use of the OncotypeDx test, which includes multiple collaborations (53). These studies were funded in part by NIH and demonstrate the value of public–private partnerships. Studies demonstrating prognostic performance, predictive performance, and cost-effectiveness of this test all have contributed to its commercial success.

Anticipate the importance of costs in future test reimbursement decisions

Regardless of whether the test will be reviewed by the FDA or offered without FDA review as an LDT, it is likely that future decision making about incorporating tests into everyday clinical medicine will become increasingly price sensitive. Sponsors and other stakeholders championing new tests should consider sponsoring cost-effectiveness studies to determine the relative value of a new test compared with the preexisting diagnostic pathway without use of the new test. Although the quality of care is paramount, cost is likely to become of increased importance in the future. Demonstrating that a new test is cost-effective is likely to be of real value in making a case for

use of a new test to all stakeholders and in particular to third-party payers (54).

Create a project plan that includes a timeline for testing and implementing the assay for demonstrating clinical utility

Test developers should consider integrating plans for demonstrating clinical utility with the analytical validation, clinical validation, and early clinical trial plans (see refs. 14 and 54, for a discussion of criteria for these stages) and consider that the evidence of clinical utility may vary over time. When appropriate, consider supplementing premarket studies with postmarket studies.

Carefully design clinical trials and analyses

Clinical studies for the evaluation of biomarkers can have three phases (55). Phase I studies are biomarker assay assessment in normal and tumor tissue and determination if a reliable and analytically valid signal can be generated. Phase II studies might be prospective designs for hypothesis generation and/or retrospective designs to examine biomarker utility; and phase III studies would demonstrate the clinical validity and utility of the biomarker in a prospective confirmatory multicenter studies (55) and/or might use modeling to combine disparate sources of data to create an indirect chain of evidence demonstrating clinical utility. Alternatively, prospective–retrospective designs may be used in phase II when adequate tissue representative of the patients in a trial is available to generate sufficient statistical power and the results can be validated using specimens from separate studies (56). A trial design for a companion diagnostic assay will depend upon whether or not the test is developed concurrently with the therapeutic or after the therapeutic has achieved FDA approval. Prominent recent examples involve adaptive randomization (BATTLE-1; ref. 57) and I-SPY 2 (58) and other enrichment designs such as the Master Protocol in Squamous Cell Lung Cancer (59). Table 5 provides a list of trial designs relevant to demonstration of utility.

In planning trials, it is suggested that developers anticipate questions and issues about clinical utility as outlined above. Focus on studies that reflect how the test will be used in clinical practice. Take into account the mechanism linking the biomarker to the disease. In sizing the trials, account for disease prevalence, appropriateness of certain designs (e.g., enrichment designs vs. all comers or stratified), as well as the ethics of drug administration and trial costs versus returns. In addition to outcomes such as improved overall survival (OS) and progression-free survival (PFS), specify outcomes and decision options in the clinical study protocol that apply to clinical utility (better patient-reported outcomes, lower resource utilization or cost, value of the test result, and other biomarker-related interventions). Review the statistical analysis plan and ensure it is appropriate for the proposed use of the new assay and demonstration of clinical utility. Include comparison to standard of care/current practice or as add-on to standard of care/
current practice and ensure recruited patients represent the intended use population. In addition, if the new test use population will be implemented in more than one setting, consider planning to perform the clinical study in 2 or more laboratories.

**Assemble the chain of evidence (study data) to demonstrate utility**

The hierarchy of clinical study evidence is important. In cancer therapeutic studies, for example, OS has traditionally been considered the gold standard in disease endpoint measurements but is not always practical, for example, when duration of likely survival is very long, when patients often cross over to other therapies, etc. FDA has approved drugs using a variety of surrogate endpoints, such as PFS, time to progression, reduction in symptoms, etc. Patient centered or patient reported outcomes such as quality of life are also of increasing interest to stakeholders in understanding disease course, treatment effect, and perceived benefit. Studies to establish clinical utility of diagnostics as well as drugs should clearly describe the relative importance of the outcomes measured. It is important to specify clearly whether or not analysis and data interpretation were blinded and whether or not negative results were independently verified. It is also important to show that the data analysis approach adequately assesses the proposed use. Either using clinical trials or using decision modeling, studies should be designed to compare results to current practice to demonstrate the value added for the new assay. This should include an assessment of the assay’s impact on physician behavior and health care.

The use of a prospective randomized controlled clinical trial studying relevant outcomes of importance to patients is the ideal method to develop a chain of evidence and allows for a direct measurement of how a test affects outcomes. However, these studies can be expensive and lengthy to perform and hence commonly are not done. Alternatively, it may be possible to acquire a high level of evidence from prospective–retrospective studies. Sometimes the only other options are to use observational studies, nonrandomized clinical trials, other types of descriptive studies, and/or information from practice guidelines to describe clinical utility in a 2-step process: (1) demonstrating a test has an impact on intermediate outcomes of interest and (2) demonstrating the outcome of interest can be shown to produce improvements in clinical outcomes. A good example of this type of approach to establishing clinical utility is the 2008 TEC-BCBSA Genetic Testing in Long QT Syndrome. At any rate, because third-party payers most often cross over to other therapies, etc. FDA has approved drugs using a variety of surrogate endpoints, such as PFS, time to progression, reduction in symptoms, etc. Patient centered or patient reported outcomes such as quality of life are also of increasing interest to stakeholders in understanding disease course, treatment effect, and perceived benefit.

By definition, regulatory requirements to receive approval for companion diagnostics include demonstration of clinical utility. Therefore, the approval in recent years of companion diagnostics developed with targeted therapies serve as the best available examples of successful clinical utility efforts. These successes include the KRAS test for EGFR antibody therapy (cetuximab and panitumumab) in metastatic colorectal cancer (60–65), EGER exon testing for EGFR kinase inhibitor treatment in metastatic non–small cell lung cancer (NSCLC; refs. 66–69), and the BRAF mutation test for mutant BRAF kinase inhibitor therapy in patients with metastatic melanoma (70). Of note, the association of wild type but not mutant KRAS with positive treatment response in colorectal and lung cancer was established by retrospective analysis using archived tissues from the efficacy trials (60–63). It should be recognized that with the ever-increasing availability of relevant data and advanced technologies new tests are being developed continually, many of which may be improvements over existing diagnostics, but lack sufficient evidence to demonstrate clinical utility. Therefore, the reality is that many patient decisions today are made on the basis of tests for which clinical utility may not be well established. This can be true for both FDA-reviewed tests and those offered without FDA review. Because the consequences, if any, of this situation are currently uncertain, collaborations meant to acquire this evidence clearly would be beneficial.

**Conclusions**

As outlined in this article and in cited references, there is growing recognition and discussion of the importance of establishing clinical utility in ensuring the successful development and clinical integration of molecular diagnostic assays. Practical translation of the factors involved in demonstrating clinical utility can be accomplished by planning from the early stages of assay formulation and development and by anticipating what types of questions will need to be answered and what kinds of data need to be generated to supplement routine analytical and clinical validation studies in demonstrating the clinical utility of a new test. Critical to this long-term plan is the ability of teams of development, regulatory, clinical, and commercialization stakeholders to work together in a timely and collaborative manner.

An assay development project plan by necessity requires early evaluation of the context of the assay. A central goal of such a plan is to examine the clinical usefulness of the test result relative to existing disease and clinical standards of care, to determine how to translate the test to a clinical laboratory setting, and to identify potential patient benefits from the test result. Early attention looking beyond CLIA or FDA regulatory review requirements in order to satisfy the interests of third-party payers is likely to make market entry smoother and more successful and to allow sponsors to plan early to ensure that the infrastructure and training are in place for effective use of the assay outside the R&D environment.

By definition, regulatory requirements to receive approval for companion diagnostics include demonstration of clinical utility. Therefore, the approval in recent years of companion diagnostics developed with targeted therapies serve as the best available examples of successful clinical utility efforts. These successes include the KRAS test for EGFR antibody therapy (cetuximab and panitumumab) in metastatic colorectal cancer (60–65), EGER exon testing for EGFR kinase inhibitor treatment in metastatic non–small cell lung cancer (NSCLC; refs. 66–69), and the BRAF mutation test for mutant BRAF kinase inhibitor therapy in patients with metastatic melanoma (70). Of note, the association of wild type but not mutant KRAS with positive treatment response in colorectal and lung cancer was established by retrospective analysis using archived tissues from the efficacy trials (60–63). It should be recognized that with the ever-increasing availability of relevant data and advanced technologies new tests are being developed continually, many of which may be improvements over existing diagnostics, but lack sufficient evidence to demonstrate clinical utility. Therefore, the reality is that many patient decisions today are made on the basis of tests for which clinical utility may not be well established. This can be true for both FDA-reviewed tests and those offered without FDA review. Because the consequences, if any, of this situation are currently uncertain, collaborations meant to acquire this evidence clearly would be beneficial.
Finally, rapid evolution of sequencing technology and genomic panels used to characterize individual patient cancers for treatment decisions presents an unresolved clinical utility question. Reference laboratories and small biotechnology companies are marketing genotyping panels for drug targets to patients, even when a therapeutic is not approved for that target/patient class. This rapidly developing technology continually provides challenges for regulatory science and ethics issues. The situation also underlines the issue of what level of evidence is needed to decide on a treatment based on genomics or histologic characteristics, without evidence that treatments determined by genomics work. These questions are also dependent on whether or not these tests or panels have established adequate analytical and clinical validation before attempting to establish clinical utility.

Given that demonstration of clinical utility is currently often a missing element in the evaluation of new laboratory tests, there is a critical need to raise the consciousness about its importance, to encourage the use of better planning and methodologies in addressing the search for utility, and to create guidance that will be of value to all stakeholders working to medical care more effective. Recognition of the importance of establishing the clinical utility of a test can be reinforced by payer resistance to tests lacking this level of evidence, or by a tiered reimbursement system which rewards clinical tests with superior levels of clinical utility information. The requirement for high levels of evidence for a developed test presents considerable challenges with respect to business models for diagnostic companies. The development of such high levels of evidence requires time, resources, and skill sets not traditionally found within diagnostic companies, and often high levels of collaboration with investigators and clinical trial sponsors. Absent a rewards system that recognizes the value of this information, complex tests of high predictive value will not be developed with the needed frequency. If the practice of clinical medicine is to benefit from the biological and technological advances resulting from the new “omics” era, regulatory and reimbursement expectations need to be more aligned. Clinical tests of high-potential clinical utility should be prioritized by payers; improved reimbursement would stimulate private investment in the

Appendix A. ACCE model list of questions for review of clinical utility for genetic tests

<table>
<thead>
<tr>
<th>Component</th>
<th>Specific question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disorder/setting</td>
<td>What is the specific clinical disorder to be studied?</td>
</tr>
<tr>
<td></td>
<td>What are the clinical findings defining this disorder?</td>
</tr>
<tr>
<td></td>
<td>What is the clinical setting in which the test is to be performed?</td>
</tr>
<tr>
<td></td>
<td>What DNA test(s) are associated with this disorder?</td>
</tr>
<tr>
<td></td>
<td>Are preliminary screening questions employed?</td>
</tr>
<tr>
<td></td>
<td>Is it a stand-alone test or is it one of a series of tests?</td>
</tr>
<tr>
<td></td>
<td>If it is part of a series of screening tests, are all tests performed in all instances (parallel) or are only some tests performed on the basis of other results (series)?</td>
</tr>
<tr>
<td>Intervention</td>
<td>What is the natural history of the disorder?</td>
</tr>
<tr>
<td></td>
<td>What is the impact of a positive (or negative) test on patient care?</td>
</tr>
<tr>
<td></td>
<td>If applicable, are diagnostic tests available?</td>
</tr>
<tr>
<td></td>
<td>Is there an effective remedy, acceptable action, or other measurable benefit?</td>
</tr>
<tr>
<td></td>
<td>Is there general access to that remedy or action?</td>
</tr>
<tr>
<td></td>
<td>Is the test being offered to a socially vulnerable population?</td>
</tr>
<tr>
<td>Quality assurance</td>
<td>What quality assurance measures are in place?</td>
</tr>
<tr>
<td>Pilot trials</td>
<td>What are the results of pilot trials?</td>
</tr>
<tr>
<td>Health risks</td>
<td>What health risks can be identified for follow-up testing and/or intervention?</td>
</tr>
<tr>
<td>Economic</td>
<td>What are the financial costs associated with testing?</td>
</tr>
<tr>
<td></td>
<td>What are the economic benefits associated with actions resulting from testing?</td>
</tr>
<tr>
<td>Facilities</td>
<td>What facilities/personnel are available or easily put in place?</td>
</tr>
<tr>
<td>Education</td>
<td>What educational materials have been developed and validated and which of these are available?</td>
</tr>
<tr>
<td>Monitoring</td>
<td>Are there informed consent requirements?</td>
</tr>
<tr>
<td></td>
<td>What methods exist for long term monitoring?</td>
</tr>
<tr>
<td></td>
<td>What guidelines have been developed for evaluating program performance?</td>
</tr>
</tbody>
</table>

Adapted from the ACCE Model List of 44 Targeted Questions Aimed at a Comprehensive Review of Genetic Testing. See the entire list (48) for questions related to analytic validity and clinical validity.
Appendix B. Components of the EGAPP methods for evaluation of clinical utility for genetic tests

<table>
<thead>
<tr>
<th>Component</th>
<th>Characteristics and questions addressed</th>
</tr>
</thead>
</table>
| Evaluation methods                                              | - Whether use of the assay improves clinical and health outcomes based on subsequent diagnosis and intervention  
                  | - Availability of information useful for decision-making  
                  | - Is this the last possible assay to inform a clinical decision?  
                  | - Balance of benefits and harms when the test results are considered in patient management  
                  | - Chain of evidence demonstrating that the test results can change patient management decisions and improve net health outcomes |
| Acceptable evidence (clinical studies)                           | - Peer-reviewed publications of data  
                  | - Review of a meta-analysis of multiple studies with the assay |
| Important study design characteristics                          | - Quality of the individual clinical studies  
                  | - Quality of the overall body of evidence  
                  | - Quality of the relevant data  
                  | - Consistency and generalizability of the findings, and whether or not implementation of tests in different settings could lead to variability in health outcomes |
| Grading system for the quality of evidence of components         | - Convincing (results from well-designed and conducted studies in representative populations that assess specified health outcomes, systematic review or meta-analysis of randomized controlled trials that show consistent results, or at least one large randomized controlled trial)  
                  | - Adequate (results that might be influenced by flawed study methodology; systematic review with heterogeneity; one or more nonrandomized, peer-reviewed, but not well-designed controlled trials; or systematic review of cohort studies with consistent results)  
                  | - Inadequate (results that are likely to come from flawed study methodology, as in systematic review of studies with heterogeneity, single-level cohort or case-control studies, or data that was neither published nor peer-reviewed)  
| Additional criteria used to assess quality                       | - Clear presentation of the study design; clear description and definitions of outcomes of interest or decision options  
                  | - Relative importance of the outcomes measured  
                  | - Blinded interpretation of outcomes  
                  | - Verification of negative results  
                  | - Use of prospective or retrospective data collection  
                  | - Appropriateness of data analysis  
                  | - Randomization of subjects in experimental designs  
                  | - Study inclusion of comparison with current practice to demonstrate value added  
                  | - Interventions used and criteria needed for the intervention  
                  | - Information on data analysis sufficient to rate the study quality  
                  | - Relevance of data to identified outcomes  
                  | - Adequate and appropriate description of the analysis or modeling for the study design and proposed use  
                  | - Consideration of losses and potential bias |
                  | - Development of methodology: S.R. Tunis, G.J. Kelloff  
                  | - Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): R.T. McCormack, C.C. Signman, G.J. Kelloff  
                  | - Study supervision: G.J. Kelloff |

Development of such tests. Additional mechanisms that would allow for testing, evaluation, and refinement of the methodologies being proposed here can include public/private activities such as those carried out by the FNIH Biomarkers Consortium.

Disclosure of Potential Conflicts of Interest

D.R. Parkinson is a CEO of Nodality, Inc. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Development of methodology: S.R. Tunis, G.J. Kelloff
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): R.T. McCormack, C.C. Signman, G.J. Kelloff
Study supervision: G.J. Kelloff
Acknowledgments
The authors wish to gratefully acknowledge the support of the Foundation for the National Institutes of Health Biomarkers Consortium in providing a venue for carrying out this project.

References
41. Hayes DF. OMICS-based personalized oncology: if it is worth doing, it is worth doing well! BMC Med 2013;11:221.
42. Meleth S, Whitehead N, Evans TS, Lux L. Technology assessment on genetic testing or molecular pathology testing of cancers with unknown primary site to determine origin. AHRQ Technology Assessment Program 2013;20:209.
52. Wooley A. FDA-approved LTD: accelerating the path to commercialization while mitigating risk and cost. Accelerating Development & Advancing Personalized Therapy Congress (ADAPT); 2012.
Evidence of Clinical Utility: An Unmet Need in Molecular Diagnostics for Patients with Cancer
